Title: Learning Implicit Models of Complex Dynamical Systems From Partial Observations

Author(s): Rupe, Adam Thomas

Intended for: Web

Issued: 2021-07-19

Learning Implicit Models of Complex Dynamical Systems From Partial Observations

Adam Rupe

(he/him)

Center for Nonlinear Studies and Computational Earth Science

Los Alamos National Laboratory

Inference For Dynamical Systems, 7/21/2021

Systems, Data, and Models

(analysis / assimilation)

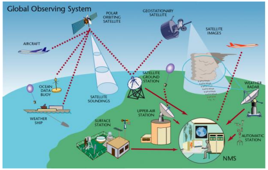
<u>System</u>

Credit: NASA–Apollo 17 crew

(measurement)

$X : \Omega \to \mathbb{R}^n$

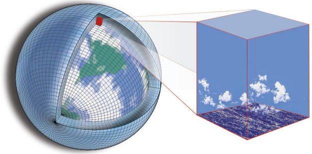
<u>Data</u>



Credit: WMO

A

<u>Data Image</u>



Credit: Tapio Schneider/Kyle Pressel/Momme Hell/Caltech

ω_0 $\qquad$ $x_0 = X(\omega_0)$ $\qquad$ $u_0 = A(x_0, x_{-1}, x_{-2}, \ldots)$

True Dynamics Φ^t

Koopman Operator $U^t X = X \circ \Phi^t$

$\mathcal{T}^t$ Data-driven model

$\widetilde{\Phi}^t_\alpha$ Physics simulation

$\widetilde{x}_t = \mathcal{T}^t(x_0)$

minimize

$\omega_t \xrightarrow{X : \Omega \to \mathbb{R}^n} x_t = X(\omega_t)$ $\qquad$ u_t

Systems and "Platonic" Models

Physical system described by measure space $(\Omega, \Sigma_\Omega, \nu)$

Governing physics captured by closed-form diff. eq. model $\dot{\omega} = \Phi(\omega)$

- *Markovian*—orbits ω_t determined by single ω_0
- *Deterministic*—orbit $\omega_t(\omega_0)$ always the same for same ω_0

closure relationship between *degrees of freedom* of ω

- ω a vector $\{\omega^i\}$—each ω^i a DoF
- evolution of single ω^i depends on interactions with other DoF
- physics "captured"/"modeled" with deterministic, Markov dynamic closed over $\{\omega^i\}$

("Unreasonable Effectiveness of Mathematics"?)

Measurements and Partial Observations

For given $(\Omega, \Sigma_\Omega, \nu, \Phi)$, collection of instrument measurements is measurable function
$X : \Omega \to \mathcal{X}, \quad \text{s. t. } x_t = X(\omega_t)$

Generally interested in *partially-observed systems*, for which X is *not invertible*: knowledge of x insufficient for determining state ω of the true system

$\implies$ "unobservable" or "immeasureable" degrees of freedom in ω not in $x = X(\omega)$

Initial observation x_0 induces uncertainty over ω_0 (e.g. MaxEnt), giving $\mu_0 : \Omega \to \mathbb{R}$

$\implies$ probability space $(\Omega, \Sigma_\Omega, \mu_t, \Phi)$, with μ_t governed by *Perron-Frobenius*

observables X are random variables governed by *Koopman*

(foundational set up for nonequilibrium statistical mechanics)

M. Mackey "Time's Arrow: The Origins of Thermodynamic Behavior", Springer (1993)

Models From Partial Observations

Explicit Models — physics simulations

- use histories $\{x_0, x_{-1}, \ldots\}$ to *explicitly* "fill in the gaps" of unobserved DoF
- fit (approximated) closed-form deterministic Markov model $\widetilde{\Phi}^{\tau}_{\alpha}$
- auxiliary dyn sys whose numerical solutions approximate evolution $\omega_t = \Phi^t(\omega_0)$

Implicit Models — data-driven methods

- No direct reference to physics (degrees of freedom or their governing equations)
- learn a mapping $\mathcal{T}^{\tau}$ that evolves measurement observables forward in time
- *implicitly* "fills in the gaps" of unobserved DoF with delay-coordinate embeddings
- learns *implicit* Markov dynamic with Koopman operator

Will show certain data-driven models do implicitly what physics simulations do explicitly

Physics-Based Simulations

data image u_t — approximated coarse-graining of system state ω_t

data assimilation — $u_0 = A(x_0, x_{-1}, \ldots)$ most consistent with history of observations

simulation $u_{t+\tau} = \widetilde{\Phi}_\alpha^\tau \, u_t$ — uses parameterization α for Markovian closure

unobserved DoF not in X *explicitly* "filled in" by coarse-graining u_t or parameterization α

$u_{t+\tau} = \widetilde{\Phi}_\alpha^\tau \, u_t$ explicitly approximates governing physics $\omega_{t+\tau} = \Phi^\tau \omega_t$

model error from parameterization α

initialization error from $u_0 = A(x_0, x_{-1}, \ldots)$ (e.g. chaos)

Instantaneous Data-Driven Models

(Instantaneous) *target function* $\mathcal{T}^\tau : \mathcal{X} \to \mathcal{X}$ learned from data

$$\text{minimize} \;\; ||\mathcal{T}^\tau \circ X - U^\tau X||^2_{L^2(\Omega,\mu_*)}$$

where $U^\tau X_t \approx x_{t+\tau}$ from training data — i.e. loss $= ||\mathcal{T}^\tau(x_t) - x_{t+\tau}||^2$

regression function Z^τ unique minimizer — aka conditional expectation $Z^\tau = \mathbb{E}[U^\tau X_t | X_t]$

Decompose model error as

$$\mathcal{E}(\mathcal{T}^\tau) := ||\mathcal{T}^\tau \circ X - U^\tau X||^2 = \Theta(\mathcal{T}^\tau) + \Xi^\tau_X$$

$\Theta(\mathcal{T}^\tau) = ||\mathcal{T}^\tau - Z^\tau||^2$ — how far given $\mathcal{T}^\tau$ is from optimum

$\Xi^\tau_X = ||Z^\tau \circ X - U^\tau X||^2$ — intrinsic error due to partial observations X

R. Alexander and D. Giannakis (2020) Physica D

Hilbert Space Formulation

System observables: $H = L^2(\mu_*)$
$H = \{f : \Omega \to \mathcal{X} : \int_\Omega ||f^2(\omega)||^2 d\mu(\omega) < \infty\}$

Functions of measurement observations:
$V = \{g : \mathcal{X} \to \mathcal{X} : g \circ X \in H\}$

Subspace of measurement observables: $H_X = L^2(\mu_*^X)$
$H_X = \{f \in H : f = g \circ X \text{ for some } g \in V\}$

Regression function the unique element in V s.t.

$$\mathbb{E}[U^\tau X_t | X_t] = Z^\tau := \operatorname*{argmin}_{g \in V} ||g \circ X - U^\tau X||_H^2$$

F. Gliani, D. Giannakis, J. Harlim (2021) Physica D

Data-Driven Approximations of U^τ

EDMD: Galerkin projection onto $H_{\boldsymbol{\Psi}} \subseteq H_X \subset H$ using $\boldsymbol{\Psi} \subseteq V$

M. O. Williams, I. G. Kevrekidis, and C. W. Rowley (2015). Journal of Nonlinear Science

S. Klus, P. Koltai, and C. Schütte (2016). Journal of Computational Dynamics

System | Observations | Galerkin Approximation

$$
\begin{tikzcd}
\omega_0 \in \Omega \arrow[r, "X : \Omega \to \mathcal{X}"] \arrow[dd, "\Phi^t"] \arrow[ddr, "X_t = U^t X"] & x_0 = X(\omega_0) \arrow[r, "\boldsymbol{\Psi}"] & {[\psi^1(x_0), \ldots, \psi^k(x_0)]^{\mathrm{T}}} \arrow[d, "\mathbf{U}_X^t"'] \\
 & & {[\mathbf{U}_X^t \psi^1(x_0), \ldots, \mathbf{U}_X^t \psi^k(x_0)]^{\mathrm{T}}} \arrow[d, dashed, leftrightarrow, "\text{minimize}"] \\
\omega_t = \Phi^t(\omega_0) \arrow[r, "X : \Omega \to \mathcal{X}"] & x_t = X(\omega_t) \arrow[r, "\boldsymbol{\Psi}"] & {[\psi^1(x_t), \ldots, \psi^k(x_t)]^{\mathrm{T}}}
\end{tikzcd}
$$

Data-Driven Approximations of U^τ

Forecast using identity function $f_X(x) = x$:

$$\mathcal{T}^\tau_{\mathrm{EDMD}}(x_t) = \mathbf{U}^\tau_X[f_X \circ X](\omega_t)$$

$\implies$ must have $f_X \in \mathbf{\Psi} \subseteq V$ express as $f_X(x) = \sum_l \eta_l \mathbf{v}_l(x)$, with $\mathbf{U}^\tau_X \mathbf{v}_l = \lambda_l \mathbf{v}_l$

Linear forecast model

$$\mathbb{E}[X_{t+\tau}|X_t] \approx \mathcal{T}^\tau_{\mathrm{EDMD}}(x_t) = \sum_l \lambda_l \eta_l \mathbf{v}_l(x_t)$$

Data-Driven Approximations of U^τ

Forecast using identity function $f_X(x) = x$:

$$\mathcal{T}^\tau_{\text{EDMD}}(x_t) = \mathbf{U}^\tau_X[f_X \circ X](\omega_t)$$

$\implies$ must have $f_X \in \mathbf{\Psi} \subseteq V$ express as $f_X(x) = \sum_l \eta_l \mathbf{v}_l(x)$, with $\mathbf{U}^\tau_X \mathbf{v}_l = \lambda_l \mathbf{v}_l$

Linear forecast model

$$\mathbb{E}[X_{t+\tau} | X_t] \approx \mathcal{T}^\tau_{\text{EDMD}}(x_t) = \sum_l \lambda_l \eta_l \mathbf{v}_l(x_t)$$

In fully-observed case ($\mathcal{X} = \Omega$): $\mathbf{U}^\tau_X \to U^\tau$ in limit $\mathbf{\Psi} \to V$ $(H_{\mathbf{\Psi}} \to H_X = H)$

M. Korda and I. Mezič (2018) Journal of Nonlinear Science

recover regression function in partially-observed case: $H_{\mathbf{\Psi}} \to H_X \subset H$ for $\mathbf{\Psi} \to V$

$$\mathbf{U}^\tau_X[f_X \circ X] = \sum_l \lambda_l \eta_l \mathbf{v} \circ X = Z^\tau \circ X$$

Rupe et. al. (2021) in progress

Intrinsic Geometry and Delay-Coordinate Embeddings

Attractor geometry of underlying system (Ω, Φ) can be recovered using *delay-coordinate embeddings* $\overleftarrow{x}^m = (x_t, x_{t-1}, \ldots, x_{t-m})$, ($m$ is embedding dimension)

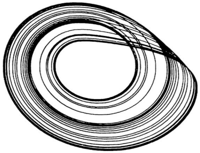


FIG. 1. (x,y) projection of Rossler (Ref. 7).

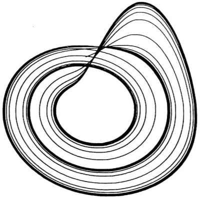


FIG. 2. $(x,\dot{x})$ reconstruction from the time series.

N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw (1980) PRL

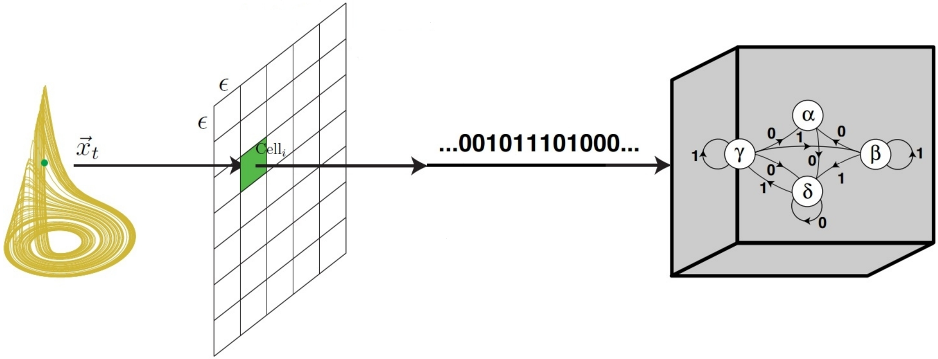
F. Takens (1981)

delay-coordinate observables: Koopman $\Longleftrightarrow$ Laplace-Beltrami operators

D. Giannakis (2019) Applied and Computational Harmonic Analysis

Intrinsic Computation of Symbolic Processes

System	Observable	Measurements	Model
$\omega_t \in \Omega$	$X : \Omega \to \mathcal{A}$	$\{x_t = X(\omega_t) \in \mathcal{A}\}$	$\overleftarrow{\mathcal{T}}^{\tau} : \overleftarrow{X}_t \mapsto X_{t+1}$



Introduced to Dynamical Systems and Ergodic theory:

Discrete Information: A.N. Kolmogorov (1959); Y.G. Sinai (1959); Y. Pesin (1977)

Discrete Computation: B. Weiss (1973); W. Krieger (1974); R. Fischer (1975); J. P. Crutchfield, K. Young (1989)

Causal States: Minimal Models for Optimal Prediction

past: $\overleftarrow{x}_t = \{\ldots, x_{t-2}, x_{t-1}, x_t\}$; future: $\overrightarrow{x}_t = \{x_{t+1}, x_{t+2}, x_{t+3}, \ldots\}$

predictive equivalence:

$$\overleftarrow{x}_i \sim_\epsilon \overleftarrow{x}_j \iff \Pr(\overrightarrow{X}|\overleftarrow{x}_i) = \Pr(\overrightarrow{X}|\overleftarrow{x}_j) \iff \epsilon(\overleftarrow{x}_i) = \epsilon(\overleftarrow{x}_j)$$

Causal States: Minimal Models for Optimal Prediction

past: $\overleftarrow{x}_t = \{\ldots, x_{t-2}, x_{t-1}, x_t\}$; future: $\overrightarrow{x}_t = \{x_{t+1}, x_{t+2}, x_{t+3}, \ldots\}$

predictive equivalence:

$$\overleftarrow{x}_i \sim_\epsilon \overleftarrow{x}_j \iff \Pr(\overrightarrow{X}|\overleftarrow{x}_i) = \Pr(\overrightarrow{X}|\overleftarrow{x}_j) \iff \epsilon(\overleftarrow{x}_i) = \epsilon(\overleftarrow{x}_j)$$

causal states: equivalence classes of $\sim_\epsilon$ generated by ϵ-function, $\epsilon : \overleftarrow{x}_t \mapsto S_t$

$$S_t = \epsilon(\overleftarrow{x}_t) = \{\overleftarrow{x}_i \ : \ \Pr(\overrightarrow{X}|\overleftarrow{x}_i) = \Pr(\overrightarrow{X}|\overleftarrow{x}_t)\}$$

Causal States: Minimal Models for Optimal Prediction

past: $\overleftarrow{x}_t = \{\ldots, x_{t-2}, x_{t-1}, x_t\}$; future: $\overrightarrow{x}_t = \{x_{t+1}, x_{t+2}, x_{t+3}, \ldots\}$

predictive equivalence:

$$\overleftarrow{x}_i \sim_\epsilon \overleftarrow{x}_j \iff \Pr(\overrightarrow{X}|\overleftarrow{x}_i) = \Pr(\overrightarrow{X}|\overleftarrow{x}_j) \iff \epsilon(\overleftarrow{x}_i) = \epsilon(\overleftarrow{x}_j)$$

causal states: equivalence classes of $\sim_\epsilon$ generated by ϵ-function, $\epsilon : \overleftarrow{x}_t \mapsto S_t$

$$S_t = \epsilon(\overleftarrow{x}_t) = \{\overleftarrow{x}_i \ : \ \Pr(\overrightarrow{X}|\overleftarrow{x}_i) = \Pr(\overrightarrow{X}|\overleftarrow{x}_t)\}$$

causal state dynamics are *Markovian*:

$$S_{t+1} = M_\epsilon S_t$$

Follows from predictive equivalence and ϵ-function: S_{t+1} uniquely determined by S_t and x_{t+1}

C. R. Shalizi and J.P. Crutchfield (2001) J. Stat. Phys.

Perspective From Mori-Zwanzig Formalism

Expand U^{t+1} in terms of *projection operator* $P : H \to H_X$ and $Q = I - P$

$$U^{t+1} = \sum_{k=0}^{t} U^{t-k} P U (QU)^k + (QU)^{t+1}$$

Apply both sides to measurement observable $X \in H$: $x_{t+1} = X_{t+1}(\omega_0) = [U^{t+1}X](\omega_0)$

Perspective From Mori-Zwanzig Formalism

Expand U^{t+1} in terms of *projection operator* $P : H \to H_X$ and $Q = I - P$

$$U^{t+1} = \sum_{k=0}^{t} U^{t-k} P U (QU)^k + (QU)^{t+1}$$

Apply both sides to measurement observable $X \in H$: $x_{t+1} = X_{t+1}(\omega_0) = [U^{t+1} X](\omega_0)$

Discrete-time MZ equation:

$$x_{t+1} = M_0(x_t) + \sum_{k=1}^{t} M_k(x_{t-k}) + \xi_{t+1}(\omega_0)$$

$M_0 \to$ Markov term $(\mathcal{T}^1)$; $\quad M_k = P(\xi_k \circ \Phi) \to$ Memory; $\quad \xi_{t+1}(\omega_0) \to$ orthogonal "noise" (Ξ_X^τ)

K. Lin and F. Lu (2020) J. Comp. Phys. F. Gliani, D. Giannakis, J. Harlim (2021) Physica D

Data-driven methods: Chorin, Hald, Kupferman (2003); Chorin, Lu (2015); Lin, Lu (2020); Lin, Tian, Livescu, Anghel (2021)

History-Dependent Models

Apply MZ expanded Koopman to delay-coordinate observables (pasts) :

$$x_{t+1} = \overleftarrow{M}_0(\overleftarrow{x}_t^m) + \text{noise}$$

Memory dependence folded into observables; noise (Ξ_X^τ ?) vanishes in $m \to \infty$ limit

K. Lin and F. Lu (2020) J. Comp. Phys. F. Gliani, D. Giannakis, J. Harlim (2021) Physica D

History-Dependent Models

Apply MZ expanded Koopman to delay-coordinate observables (pasts) :

$$x_{t+1} = \overleftarrow{M}_0(\overleftarrow{x}_t^m) + \text{noise}$$

Memory dependence folded into observables; noise (Ξ_X^τ ?) vanishes in $m \to \infty$ limit

K. Lin and F. Lu (2020) J. Comp. Phys. F. Gliani, D. Giannakis, J. Harlim (2021) Physica D

History-dependent data-driven models: $\overleftarrow{T}_\tau^m : \overleftarrow{X}_t^m \mapsto X_{t+\tau}$ $(\overleftarrow{T}_1 = \overleftarrow{M}_0)$

$$\text{minimize } ||\overleftarrow{T}_\tau^m \circ \overleftarrow{X}^m - U^\tau \circ X||^2$$

History-Dependent Models

Apply MZ expanded Koopman to delay-coordinate observables (pasts) :

$$x_{t+1} = \overleftarrow{M}_0(\overleftarrow{x}_t^m) + \text{noise}$$

Memory dependence folded into observables; noise (Ξ_X^τ ?) vanishes in $m \to \infty$ limit

K. Lin and F. Lu (2020) J. Comp. Phys. F. Gliani, D. Giannakis, J. Harlim (2021) Physica D

History-dependent data-driven models: $\overleftarrow{T}_\tau^m : \overleftarrow{X}_t^m \mapsto X_{t+\tau}$ $(\overleftarrow{T}_1 = \overleftarrow{M}_0)$

$$\text{minimize } ||\overleftarrow{T}_\tau^m \circ \overleftarrow{X}^m - U^\tau \circ X||^2$$

Linear approximations of $\overleftarrow{M}_0$ using Hankel matrix methods $\begin{bmatrix} | & & | \\ \overleftarrow{x}_t^m & \ldots & \overleftarrow{x}_{t+N}^m \\ | & & | \end{bmatrix}$

Brunton, Brunton, Proctor, Kaiser, Kutz (2017) Nature Comms; Arbabi, Mezic (2017) SIAM App. Dyn. Sys.

Intrinsic Geometry Meets Intrinsic Computation

Optimize for *all* lead times τ $\implies$ reconstruct predictive distributions $\Pr(\overrightarrow{X} | \overleftarrow{x})$

Intrinsic Geometry Meets Intrinsic Computation

Optimize for *all* lead times τ $\implies$ reconstruct predictive distributions $\Pr(\overrightarrow{X}|\overleftarrow{x})$

unifilarity /causality: one-to-one corresp. between $\overleftarrow{M}_0$ (intrinsic geometry) and M_ϵ (intrinsic computation)

Theorem: Action of $\overleftarrow{M}_0$ always same for every $\overleftarrow{x}_t \in S_t$; and

Output of $\overleftarrow{M}_0$ (realization x_{t+1}) "on S_t" uniquely determines output of M_ϵ (S_{t+1})

by definition: $\Pr(X_{t+1}|\overleftarrow{x}_t) = \Pr(X_{t+1}|S_t = \epsilon(\overleftarrow{x}_t))$,

$\Pr(X_{t+1}|\overleftarrow{x}_t)$ determined by $\overleftarrow{M}_0$, and $\epsilon(\overleftarrow{x}_i) = \epsilon(\overleftarrow{x}_j) \implies \overleftarrow{M}_0(\overleftarrow{x}_i) = \overleftarrow{M}_0(\overleftarrow{x}_j)$

$S_t = \epsilon(\overleftarrow{x}_t)$ and x_{t+1} uniquely determines $S_{t+1} = \epsilon(\overleftarrow{x}_{t+1}) = \epsilon(\overleftarrow{x}_t x_{t+1})$ $\square$

Rupe et. al. (2021) in progress

Connecting Explicit and Implicit Models

(* disregarding noise / error*)

Implicit Data-Driven Models

$$x_{t+\tau} = \overleftarrow{M}_0^{\tau}(\overleftarrow{x}_t^m)$$

Explicit Physics Simulation Models

$$x_{t+\tau} = \widetilde{M}_0^{\tau}(\overleftarrow{x}_t^m)$$

where $\widetilde{M}_0 = \widetilde{X} \circ \widetilde{\Phi}_{\alpha}^{\tau} \circ A$, $\quad \widetilde{X}$ "simulation measurements"

Rupe et. al. (2021) in progress

Partitioning the Past

Continuous Histories Assumption

$$\overleftarrow{x}_i \;\rightarrow\; \overleftarrow{x}_j \implies \Pr(\overrightarrow{X}|\overleftarrow{x}_i) \rightarrow \Pr(\overrightarrow{X}|\overleftarrow{x}_j)$$

G.M. Goerg and C.R. Shalizi (2012)

Equivalently: ϵ-function *absolutely continuous* over $\overleftarrow{\mathcal{X}}$

Partitioning the Past

Continuous Histories Assumption

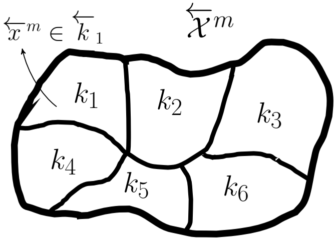
$$\overleftarrow{x}_i \;\to\; \overleftarrow{x}_j \implies \Pr(\overrightarrow{X} | \overleftarrow{x}_i) \to \Pr(\overrightarrow{X} | \overleftarrow{x}_j)$$

G.M. Goerg and C.R. Shalizi (2012)

Equivalently: ϵ-function *absolutely continuous* over $\overleftarrow{\mathcal{X}}$

Operational, approximate version—based on γ-partition:

$$\gamma(\overleftarrow{x}_i^m) = \gamma(\overleftarrow{x}_j^m) \implies \epsilon(\overleftarrow{x}_i^m) = \epsilon(\overleftarrow{x}_j^m)$$



Approximated ϵ-function *piece-wise constant* — converges* in ∞-limit

Rupe et. al. (2019) MLHPC   Rupe et. al. (2021) in progress

Causal State Analog Forecasting Algorithm

- collect finite-length pasts, futures from training data
- cluster pasts into K γ-partition elements $\mathbf{K} = \{\overleftarrow{k}_1, \overleftarrow{k}_2, \ldots, \overleftarrow{k}_K\}$ using K-Means
- for each $\overleftarrow{k}_i \in \mathbf{K}$, collect associated futures $\{\overrightarrow{x}_t : \overleftarrow{x}_t \in \overleftarrow{k}_i\}$
- this gives *sample measure* $\overrightarrow{\mu}_n = \frac{1}{n}\sum_{j=1}^{n} \delta_{\overrightarrow{x}_j}$
- construct kernel density estimator: $\widehat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} K_h(\overrightarrow{x}, \overrightarrow{x}_j)\, \delta_{\overrightarrow{x}_j}$

Predictive distributions approximated as:

$$\Pr(\overrightarrow{X} | \overleftarrow{X} = \overleftarrow{x}) \sim \frac{1}{n}\sum_{i=1}^{n} c_i K_h(\overrightarrow{x}, \overrightarrow{x}_i)\, \delta_{\overrightarrow{x}_j}$$

For CSAF have:

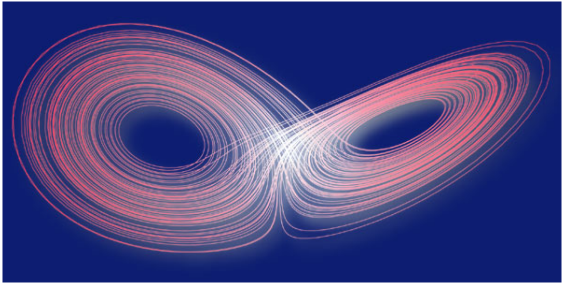
$$c_i = \begin{cases} 1 \text{ if } \gamma(\overleftarrow{x}_i) = \gamma(\overleftarrow{x}) \\ 0 \text{ otherwise} \end{cases}$$

For CONCAUST have:

$$c(\overleftarrow{x}) = (G^{\overleftarrow{X}} + \epsilon I)^{-1} K(\overleftarrow{x})$$

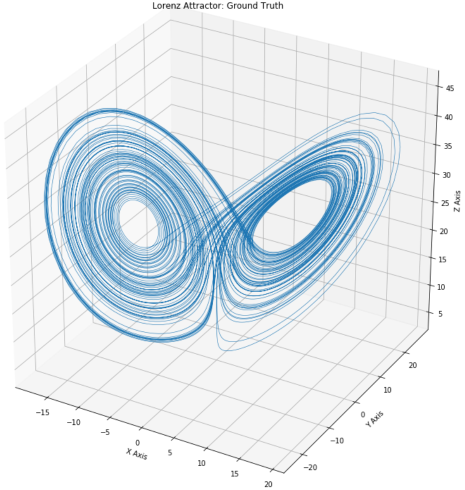
Forecasting the Lorenz Model

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x)$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z$$

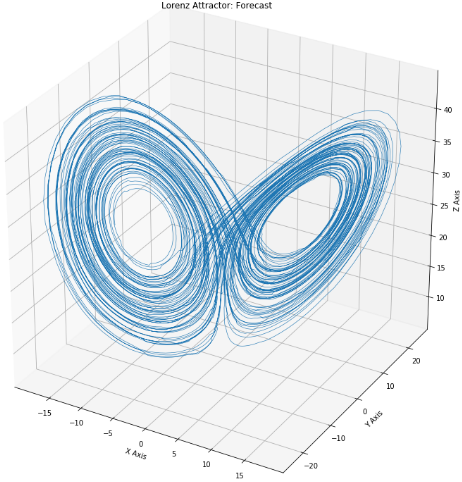


Fully-Observed Case

Numerical evolution of (x, y, z)

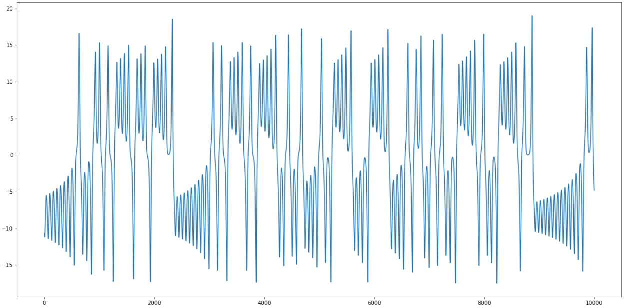


CSAF forecast of (x, y, z)

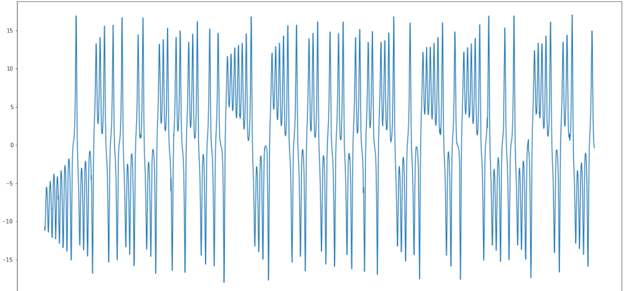


Partially-Observed Case (x variable only)

Numerical evolution of x



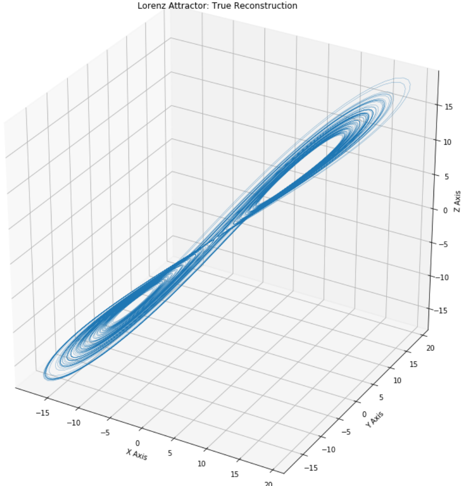
CSAF forecast of x

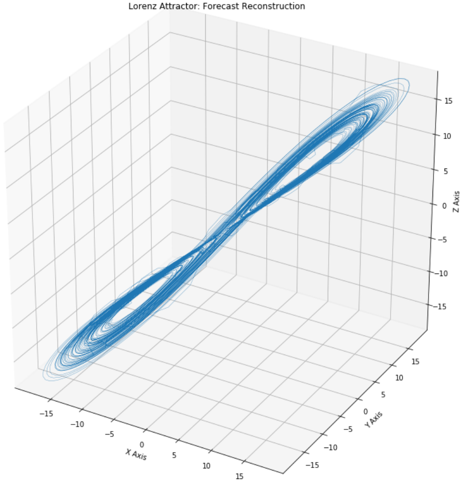


Delay-Coordinate Reconstructions

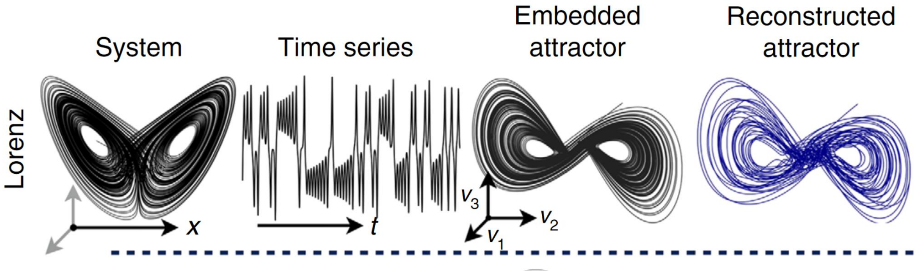
Reconstruction from true evolution (x_t, x_{t-1}, x_{t-2})

CSAF forecast reconstruction (x_t, x_{t-1}, x_{t-2})





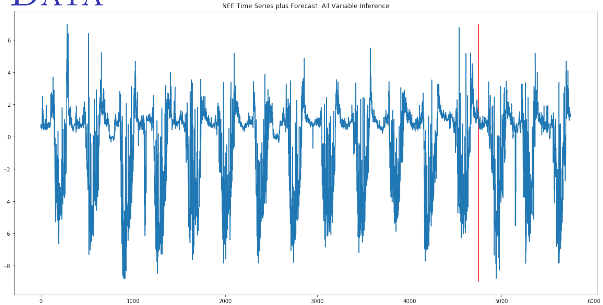
Hankel DMD / HAVOK Lorenz Forecast



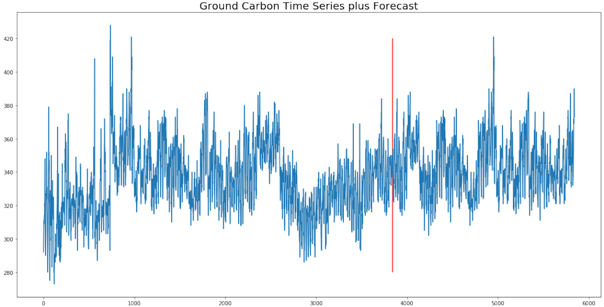
S.L. Brunton, B.W. Brunton, J.L. Proctor, E. Kaiser, J.N. Kutz
Chaos as an intermittently forced linear system
Nature Comms (2017)

Real-World Carbon Data

Training data and forecast of net ecological exchange (NEE)

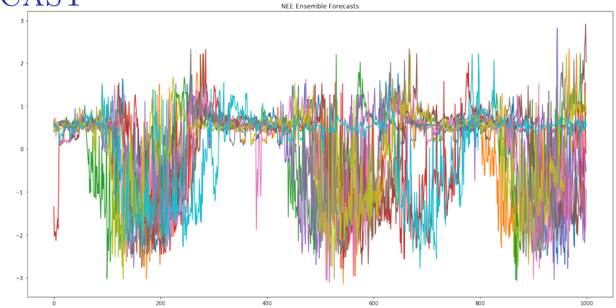


Training data and forecast of ground carbon

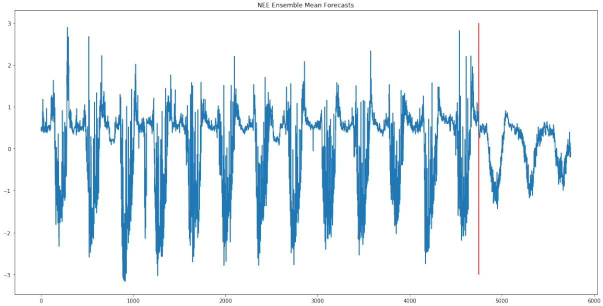


NEE Ensemble Forecast

10 ensemble members of NEE forecast



Mean forecast of 20-member ensemble NEE forecast



Conclusions

- certain data-driven models do implicitly what physics simulations do explicitly
- theoretical underpinnings in Koopman theory, delay-coordinate embeddings, and Mori-Zwanzig formalism
- best approach for given system with finite data an open question
- physics-informed machine learning (PIML) emerging framework for combining explicit and implicit modeling

Thanks to: Monty Vesselinov, Aric Hagberg, Dan O'Malley, Jim Crutchfield, Nicolas Brodu, Derek DeSantis, Stefan Klus, Balu Nadiga, Paul Riechers, Mitko Vassilev

Thank you!

Kernel Analog Forecasting

training data: data $= \{x_0, x_1, \ldots, x_T\}$

(regular) *Analog Forecasting*: identify analog $x_a \in$ data via $a = \underset{i \in \{0,\ldots,T-\tau\}}{\operatorname{argmin}} D(x, x_i)$

Forecast of x is what follows time τ after the analog

$$\mathcal{T}^{\tau}_{\mathrm{AF}}(x) = x_{a+\tau} = \int_{\mathcal{X}} U^{\tau} X_t \, d\delta_{x_a}$$

Kernel Analog Forecasting

training data: data $= \{x_0, x_1, \ldots, x_T\}$

(regular) *Analog Forecasting*: identify analog $x_a \in$ data via $a = \underset{i \in \{0,\ldots,T-\tau\}}{\operatorname{argmin}} D(x, x_i)$

Forecast of x is what follows time τ after the analog

$$\mathcal{T}^{\tau}_{\mathrm{AF}}(x) = x_{a+\tau} = \int_{\mathcal{X}} U^{\tau} X_t \, d\delta_{x_a}$$

Kernel Analog Forecasting: use similarity kernel to evaluate ensemble of analogs

$\mathcal{T}^{\tau}_{\mathrm{KAF}}(x) = \int_{\mathcal{X}} U^{\tau} X_t \, d\widehat{\mu}_X$ with $\widehat{\mu}_X = \frac{1}{n} \sum_{i=1}^{n} K_h(x, x_i) \delta_{x_i}$

$$\Pr(X_{t+\tau} | X_t = x) \sim \frac{1}{n} \sum_{i=1}^{n} K_h(x, x_i) x_{i+\tau} \, \delta_{x_i}$$

Z. Zhao and D. Giannakis (2016) Nonlinearity